ENERGY-VS-PERFORMANCE TRADE-OFFS IN SPEECH ENHANCEMENT IN WIRELESS ACOUSTIC SENSOR NETWORKS

Fernando de la Hucha Arce[1], *Fernando Rosas*[2], *Marc Moonen*[1], *Marian Verhelst*[2], *Alexander Bertrand*[1]

KU Leuven, Dept. of Electrical Engineering (ESAT), STADIUS[1], MICAS[2]
Kasteelpark Arenberg 10, 3001 Leuven, Belgium

Email: {*fernando.delahuchaarce, fernando.rosas, marc.moonen, marian.verhelst, alexander.bertrand*}*@esat.kuleuven.be*

ABSTRACT

Distributed algorithms allow wireless acoustic sensor networks (WASNs) to divide the computational load of signal processing tasks, such as speech enhancement, among the sensor nodes. However, current algorithms focus on performance optimality, oblivious to the energy constraints that battery-powered sensor nodes usually face. To extend the lifetime of the network, nodes should be able to dynamically scale down their energy consumption when decreases in performance are tolerated. In this paper we study the relationship between energy and performance in the DANSE algorithm applied to speech enhancement. We propose two strategies that introduce flexibility to adjust the energy consumption and the desired performance. To analyze the impact of these strategies we combine an energy model with simulations. Results show that the energy consumption can be substantially reduced depending on the tolerated decrease in performance. This shows significant potential for extending the network lifetime using dynamic system reconfiguration.

Index Terms— Dynamic system reconfiguration, distributed signal processing, wireless acoustic sensor networks

1. INTRODUCTION

Speech enhancement is a field in audio signal processing where the goal is to improve the quality and/or intelligibility of a speech signal corrupted by noise. The need to enhance a speech signal arises in several applications such as speech communication and speech recognition, hearing aids, computer games, etc. In order to exploit spatial diversity, several microphone arrays equipped with wireless communication capabilities can be deployed, enabling them to cooperate by exchanging processed signals to jointly execute a given signal processing task. In this way, each array has access to more audio signals captured at different locations. The resulting system is referred to as a wireless acoustic sensor network (WASN), which we define as a collection of battery-powered sensor nodes, distributed over an area of interest, where each node is equipped with several microphones, a processing unit and a wireless communications module.

In WASNs, distributed algorithms are preferred due to their ability to divide the computational effort among the sensor nodes. However, optimizing the data exchange among nodes becomes a crucial matter due to the high energy cost of wireless communications, even when using low-power technology [1]. The distributed adaptive node-specific signal estimation (DANSE) algorithm has been proven to converge to the centralized linear minimum mean squared error (MMSE) estimator with reduced data exchange in [2, 3], and has been applied to speech enhancement [4]. Nevertheless, the focus on performance optimality may lead to short network lifetime, since the algorithm requires frequent communication and is executed with fixed parameters, such as the number of active nodes or the bandwidth and bit resolution of the exchanged signals. Adjusting these parameters allows nodes to reduce their energy consumption at the cost of reduced performance, resulting in an energy-vs-performance (EvP) trade-off. To extend the lifetime of the network while keeping a reasonable performance, it is necessary that nodes exploit this trade-off to wisely invest the available energy.

In this paper, we study the influence of the aforesaid parameters on the performance of DANSE and on the energy consumption of each node in a WASN. We explain the EvP trade-offs associated with reducing the bandwidth and bit resolution of the exchanged signals, and how they add flexibility to scale the energy consumption and the speech enhancement performance. To analyze the impact of these strategies we combine an energy model with simulations. The results show that the energy consumption can be significantly reduced depending on the tolerated impact on performance. Besides, they show potential for dynamic network and node reconfigurability as a function of the performance requirements and network lifetime.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of Research Project FWO nr. G.0763.12 'Wireless Acoustic Sensor Networks for Extended Auditory Communication', Research Project FWO nr. G.0931.14 'Design of distributed signal processing algorithms and scalable hardware platforms for energy-vs-performance adaptive wireless acoustic sensor networks', and the FP7-ICT FET-Open Project Heterogeneous Ad-hoc Networks for Distributed, Cooperative and Adaptive Multimedia Signal Processing (HANDiCAMS)', funded by the European Commission under Grant Agreement no. 323944. The scientific responsibility is assumed by its authors.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

2. SIGNAL MODEL AND THE DANSE ALGORITHM

2.1. Signal model

We consider a WASN composed of K nodes, where the k-th node has access to M_k microphones. We denote the set of nodes by $\mathcal{K} = \{1, \dots, K\}$ and the total number of microphones by $M = \sum_{k \in \mathcal{K}} M_k$. The signal y_{km} captured by the m-th microphone of the k-th node can be described in the frequency domain as

$$y_{km}(\omega) = x_{km}(\omega) + v_{km}(\omega), \quad m \in \{1 \dots M_k\}, \tag{1}$$

where $x_{km}(\omega)$ is the desired speech signal component and $v_{km}(\omega)$ is the undesired noise component. In a pratical setting, each signal is processed in frames of length L, on which an L-point discrete Fourier transform (DFT) is applied (see Section 2.3). Each sample in the frame is encoded with B bits.

We denote by $\mathbf{y}_k(\omega)$ the $M_k \times 1$ vector whose elements are the signals $y_{km}(\omega)$ of node k, and $\mathbf{y}(\omega)$ as the $M \times 1$ vector in which all $\mathbf{y}_k(\omega)$ are stacked. The vectors $\mathbf{x}_k(\omega)$, $\mathbf{v}_k(\omega)$, $\mathbf{x}(\omega)$ and $\mathbf{v}(\omega)$ are defined in a similar manner. Throughout this paper, we assume that there is a single[1] desired speech source $s(\omega)$. The desired speech signal components are then given by

$$\mathbf{x}_k(\omega) = \mathbf{a}_k(\omega) s(\omega), \quad \forall k \in \mathcal{K}, \tag{2}$$

where $\mathbf{a}_k(\omega)$ is an $M_k \times 1$ vector containing the acoustic transfer functions from the source to each microphone, including room acoustics and microphone characteristics.

2.2. The DANSE algorithm

In a speech enhancement application in a WASN, the goal of the k-th node is to obtain an estimate of the speech signal component captured by one of its microphones, for instance the first microphone signal $x_{k1}(\omega)$. The linear MMSE estimator $\hat{\mathbf{w}}_k$ is given by

$$\hat{\mathbf{w}}_k = \arg\min_{\mathbf{w}_k} E\left\{ |x_{k1} - \mathbf{w}_k^H \mathbf{y}|^2 \right\}, \tag{3}$$

where $E\{\cdot\}$ is the expectation operator and the superscript H denotes conjugate transpose. For conciseness, we omit the variable ω from now on, but we note that (3) has to be solved for each frequency ω. The solution to (3) is known as multi-channel Wiener filter (MWF), and is given by [2]

$$\hat{\mathbf{w}}_k = \mathbf{R}_{yy}^{-1} \mathbf{R}_{xx} \, \mathbf{e}_1, \tag{4}$$

where $\mathbf{R}_{yy} = E\{\mathbf{y}\mathbf{y}^H\}$, $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\}$ and $\mathbf{e}_1$ is the $M \times 1$ vector $\mathbf{e}_1 = [1, 0, 0, \dots, 0]^T$. A key drawback of solving (3) in a WASN is that it requires the node to have access to $\mathbf{y}$. This means that all microphone signals y_{km} have to be exchanged between the nodes, which is unaffordable for battery-powered nodes.

[1] We note here that the DANSE algorithm can handle any number of desired sources [2, 3], but we use this assumption to simplify our EvP analysis.

The DANSE algorithm finds the node-specific estimated signals $\{\hat{\mathbf{w}}_k^H \mathbf{y}, \forall k \in \mathcal{K}\}$ without the need to exchange all the microphone signals $\mathbf{y}_k$ [2, 3]. We consider a fully connected network as it is the simplest case, but we note that the algorithm has also been adapted for a network with a tree topology [5]. The main idea of the DANSE algorithm is that each node broadcasts a linearly compressed single-channel signal

$$z_k = \mathbf{f}_k^H \mathbf{y}_k, \quad \forall k \in \mathcal{K}, \tag{5}$$

which every other node can receive. The compression filter $\mathbf{f}_k$ will be defined later (see (10)). The $K \times 1$ vector collecting all broadcast signals is denoted by $\mathbf{z} = [z_1, \dots, z_K]^T$. Each node has now access to $\tilde{M}_k = M_k + K - 1$ signals, which are stacked in the vector

$$\tilde{\mathbf{y}}_k = \begin{bmatrix} \mathbf{y}_k \\ \mathbf{z}_{-k} \end{bmatrix}, \tag{6}$$

where $\mathbf{z}_{-k}$ denotes the vector $\mathbf{z}$ with the entry z_k removed. The vectors $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{v}}_k$ are similarly defined. Then, each node computes an MWF $\tilde{\mathbf{w}}_k$ given by [2]

$$\tilde{\mathbf{w}}_k = \mathbf{R}_{\tilde{y}_k \tilde{y}_k}^{-1} \mathbf{R}_{\tilde{x}_k \tilde{x}_k} \tilde{\mathbf{e}}_1, \tag{7}$$

where $\mathbf{R}_{\tilde{y}_k \tilde{y}_k} = E\{\tilde{\mathbf{y}}_k \tilde{\mathbf{y}}_k^H\}$, $\mathbf{R}_{\tilde{x}_k \tilde{x}_k} = E\{\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^H\}$. and $\tilde{\mathbf{e}}_1$ is the $\tilde{M}_k \times 1$ vector $\tilde{\mathbf{e}}_1 = [1, 0, 0, \dots, 0]^T$. We can partition $\tilde{\mathbf{w}}_k$ in two multi-channel filters, one applied to $\mathbf{y}_k$ and one applied to $\mathbf{z}_{-k}$, as follows:

$$\tilde{\mathbf{w}}_k = \begin{bmatrix} \mathbf{h}_k \\ \mathbf{g}_k \end{bmatrix}, \tag{8}$$

and write the estimated speech component at the k-th node as

$$\hat{x}_{k1} = \tilde{\mathbf{w}}_k^H \tilde{\mathbf{y}} = \mathbf{h}_k^H \mathbf{y}_k + \mathbf{g}_k^H \mathbf{z}_{-k}. \tag{9}$$

In the DANSE algorithm, the compression filter in (5) is

$$\mathbf{f}_k = \mathbf{h}_k, \quad \forall k \in \mathcal{K}. \tag{10}$$

Notice that $\mathbf{h}_k$ is also part of the estimator in (7). However, the computation of (7) relies on access to the compressed signals $\mathbf{z}_{-k}$. To solve this problem, the set $\{\mathbf{h}_k, \forall k \in \mathcal{K}\}$ is initialized with random vectors, and then every node follows an iterative process where $\tilde{\mathbf{w}}_k$ and $\mathbf{f}_k$ are updated according to (7)-(10), based on the most recent values of $\tilde{\mathbf{y}}_k$.

Under assumption (2), it is proven in [2, 3] that the set $\{\tilde{\mathbf{w}}_k, \forall k \in \mathcal{K}\}$ converges to a stable equilibrium where, at each node k, the estimated signal in (9) is equal to the centralized node-specific MWF output signal $\hat{\mathbf{w}}_k^H \mathbf{y}$.

2.3. Implementation details

For the EvP study we focus on DANSE with simultaneous updates, named rS-DANSE, since it provides faster convergence [3]. The algorithm is implemented in a weighted overlap-add framework, in the same way as [4], using a root-Hann window with 50% overlap. This procedure allows to select the

frame length L equal to the DFT length and, as the audio signals are real, the filters $\tilde{\mathbf{w}}_k$ are estimated at the frequencies $\{\omega_l = 2\pi\frac{l}{L}, l \in \{0,\dots,L/2\}\}$. Since the speech components at the k-th node $\tilde{\mathbf{x}}_k$ are not observable, the correlation matrix $\mathbf{R}_{\tilde{x}_k\tilde{x}_k}$ cannot be estimated using temporal averaging. However, due to the independence of $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{v}}_k$, it can be estimated as $\mathbf{R}_{\tilde{x}\tilde{x}} = \mathbf{R}_{\tilde{y}_k\tilde{y}_k} - \mathbf{R}_{\tilde{v}_k\tilde{v}_k}$. The noise correlation matrix $\mathbf{R}_{\tilde{v}_k\tilde{v}_k} = E\{\tilde{\mathbf{v}}_k\tilde{\mathbf{v}}_k^H\}$ can be estimated during silence periods, when the desired speech source is not active. A voice activity detection (VAD) module is necessary to use this strategy. The correlation matrices $\mathbf{R}_{\tilde{y}_k\tilde{y}_k}$ and $\mathbf{R}_{\tilde{v}_k\tilde{v}_k}$ are estimated using a forgetting factor $0 \ll \lambda < 1$. Since the statistics of the compressed signals $\mathbf{z}$ change with each update, a sufficient number of new frames is needed to achieve a reliable estimation of the correlation matrices. The parameter N_{min} sets the minimum number of frames of 'speech and noise' and 'noise' that have to be collected before an update is performed.

3. ENERGY VS PERFORMANCE TRADE-OFFS

A straightforward strategy to extend the lifetime of the network is to reduce the number of active nodes. However, shutting down nodes can have a too large impact on the speech enhancement performance.

Since the communication costs are orders of magnitude higher than the computation costs, is interesting to explore more flexible options which keep the nodes active but reduce the amount of data they need to exchange. Therefore, in this section we propose two strategies for achieving a more flexible EvP trade-off: reducing the bandwidth and the bit resolution of the shared signals $\mathbf{z}$.

3.1. Shared bandwidth reduction

Until now, we have considered distributed speech enhancement over the whole available speech bandwidth, which is half of the sampling frequency f_s used by the nodes. In order to obtain the optimal multi-channel filter (7), every node has to transmit the complete set of DFT coefficients of its compressed signal $\{z_k(\omega_l), \forall l \in \{0,\dots,L/2\}\}$. However, if we relax our optimality goal for the whole bandwidth, nodes can compute (7) only at certain frequencies. At the remaining frequencies, nodes can compute a local MWF based only on their own microphone signals, given by

$$\mathbf{w}_k^{\text{local}} = \mathbf{R}_{y_k y_k}^{-1}\mathbf{R}_{x_k x_k}\mathbf{e}_1, \tag{11}$$

where $\mathbf{R}_{y_k y_k} = E\{\mathbf{y}_k\mathbf{y}_k^H\}$ and $\mathbf{R}_{x_k x_k} = E\{\mathbf{x}_k\mathbf{x}_k^H\}$. Notice that this divides the bandwidth in the part where spatial information from other nodes is used and the part where the node relies only on its own spatial information.

We can look at the effects of this modification from the perspectives of performance reduction and energy saving. In terms of enhancement performance, low frequencies (below 1 kHz) are more important for speech perception [6]. This suggests the use of distributed enhancement for low frequencies and local enhancement for high frequencies to ensure a smooth decrease in performance. We denote by L_{sh} the index of the maximum frequency $\omega_{L_{\text{sh}}}$ where (7) is computed.

In terms of energy saving, nodes only need to share L_{sh} DFT coefficients instead of $L/2+1$. The communication cost grows with the number of coefficients transmitted, and thus reducing the shared bandwidth allows nodes to reduce their energy consumption. Besides, notice that the local estimator (11) involves $M_k \times M_k$ matrices, which are smaller than the $\tilde{M}_k \times \tilde{M}_k$ matrices required in (7). This means that the computational cost also decreases when using shared bandwidth reduction, as we explain in Section 4.1.

3.2. Quantization of shared signals

Another way to reduce the energy spent in communication is to use less bits to quantize the DFT coefficients of the broadcast signals $z_k(\omega_l)$, thereby reducing the number of bits that need to be transmitted. The quantization of a real number $a \in [-A/2, A/2]$ with Q bits can be expressed as

$$\breve{a} = \Delta \left\lfloor \frac{|a|}{\Delta} + \frac{1}{2} \right\rfloor \text{sgn}(a), \tag{12}$$

where $\Delta = A/2^Q$ and $\text{sgn}(\cdot)$ is the signum function. As mentioned in Section 2.1, nodes executing the rS-DANSE algorithm use B bits to encode a signal sample for processing, but in order to save energy they can apply (12) with $Q < B$ bits to the real and imaginary parts of $z_k(\omega_l)$ before transmission. In terms of performance, the effect of this modification is to add an additional error to the signal estimate (9).

4. ENERGY MODEL

4.1. Computational cost

We use the term 'computational cost' for the energy spent by a node in performing the operations specified by the rS-DANSE algorithm, including the modifications described in Section 3. These operations are additions and multiplications, and are measured in floating-point operations (flops). In order to count the required flops, we have divided the processing tasks of each node per new audio frame in four steps:

1. Acquire and compress the signal frames
2. Update the correlation matrices
3. Update the filters
4. Estimate the desired speech signal frame.

We have summarized in Table 1 the number of flops required by each step for each audio frame of length L. The variable $\tilde{M}_k$ was defined in Section 2.2. The cost of performing an FFT is taken to be $5L\log_2 L$ flops. To convert from the number of flops to energy consumption, we assume that every flop consumes the same energy E_{flop}, which is determined by the hardware executing the algorithm. We have neglected

Step	Number of operations
1	$M_k\ (L+5L\log_2 L)+(2M_k-1)(L_{\text{sh}}+1)$
2	$4\tilde{M}_k^2(L_{\text{sh}}+1)+4M_k^2(L/2-L_{\text{sh}})$
3	$(\frac{1}{3}\tilde{M}_k^3+2\tilde{M}_k^2)(L_{\text{sh}}+1)+(\frac{1}{3}M_k^3+2M_k^2)(L/2-L_{\text{sh}})$
4	$(2\tilde{M}_k-1)(L_{\text{sh}}+1)+(2M_k-1)(L/2-L_{\text{sh}})+5L\log_2 L+L$

Table 1. Operations per new signal frame in rS-DANSE$_1$

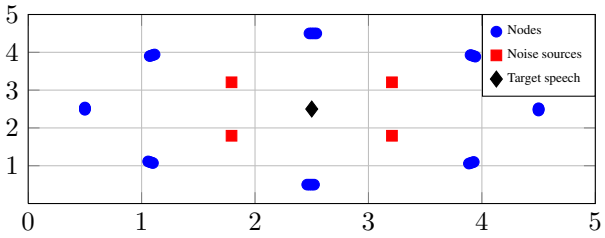

Fig. 1. Schematic of the acoustic scenario.

the cost associated with memory access, making our computational cost model optimistic.

We notice that step 3 is the most costly step. However, as opposed to steps 1, 2 and 4, this step is not performed for every new frame, but only when a sufficient number N_{min} of 'speech' and 'noise' frames have been collected to achieve a reliable estimation of the correlation matrices. A low value yields better tracking, but increases the computational cost and yields larger estimation errors in the correlation matrices.

4.2. Communication cost

For every new audio frame, the rS-DANSE algorithm requires each node to broadcast one DFT frame of size L_{sh} and to receive $K-1$ frames from the other nodes. Therefore, the communication cost for each node per audio frame is given by

$$E_{\text{comm}} = 2\,Q\,L_{\text{sh}}\left(E_{\text{cbit}}^{\text{tx}} + (K-1)E_{\text{cbit}}^{\text{rx}}\right)\ , \qquad (13)$$

where Q is the number of bits used to encode $z_k(\omega_l)$, and the factor 2 accounts for each coefficient being a complex number. The variables $E_{\text{cbit}}^{\text{tx}}$ and $E_{\text{cbit}}^{\text{tx}}$ are the energy spent to succesfully transmit and receive one bit. It includes the energy spent by the electronics of the transmitter, the radiation of the electromagnetic signal, the costs of acknowledgement signals and possible retransmissions. Due to the behaviour of wave propagation, $E_{\text{cbit}}^{\text{tx}}$ and $E_{\text{cbit}}^{\text{rx}}$ are random variables which depend on the SNR observed at the receiver. We use the analysis done in [7] to characterize the average of these quantities.

5. SIMULATION RESULTS

In order to illustrate the EvP trade-offs we explained in Section 3, we have simulated a WASN in the acoustic scenario represented in Fig 1. It consists of a cubic room of dimensions $5\times5\times5$ m, with a reverberation time of 0.2 s. In the room there are four babble noise sources and a desired speech source. All sources are located at a height of 1.8 m. The desired speech signal is a concatenation of sentences from the TIMIT database and periods of silence, with a total duration of 140.73 s. The WASN consists of eight nodes, placed 2.5 m high, where each node is equipped with 4 omnidirectional microphones. The inter-microphone distance at each node is 2 cm and the sampling rate is 16 kHz. The broadband input SNR for every node lies between -2.7 dB and -2 dB. The acoustics of the room are modeled using a room impulse response generator, which allows to simulate the impulse response between a source and a microphone using the image method. The code is available online[2]. In all simulations, we use a DFT length $L = 512$, a forgetting factor $\lambda = 0.995$ and N_{min} is set to 188, which is the number of frames collected in 3 seconds. An ideal VAD is used to exclude the influence of speech detection errors. The energy parameters of the nodes are selected to be $E_{\text{flop}} = 1$ nJ, $E_{\text{cbit}}^{\text{tx}} = 100$ nJ and $E_{\text{cbit}}^{\text{rx}} = 100$ nJ. These values represent sensor nodes, such as Zigduino [8], which use a radio compatible with the IEEE 802.15.4 standard.

In order to assess the speech enhancement performance we focus on two aspects; the noise reduction achieved and the speech distortion introduced by the filtering.

5.1. Noise reduction performance

In order to evaluate the noise reduction performance, we chose as a measure the speech intelligibility (SI) weighted SNR, where the speech and noise signals are filtered separately by one-third octave bandpass filters, and the SNR is computed per band. The SI-weighted SNR gain is defined as

$$\Delta\text{SNR}_{\text{SI}} = \sum_i I_i(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}), \qquad (14)$$

where the weight I_i expresses the importance for intelligibility of the i-th one-third octave band with center frequency $f_{c,i}$. The values for $f_{c,i}$ and I_i are defined in [9].

The SI-weighted SNR improvement is plotted as a function of the energy spent by each node in Fig. 2. Each curve in the figure corresponds to a particular choice of L_{sh} and Q, and the different marks indicate the number of active nodes (e.g. the first mark of each curve indicates one active node, and the last mark indicates eight active nodes). We define the shared bandwidth reduction parameter as $b_{\text{sh}} = L_{\text{sh}}/(L/2)$. We observe, for instance comparing the circle and square marks for the same number of nodes, that decreasing Q up to 6 bits yields a moderate reduction in performance, while the energy consumption is up to one third of the energy consumed when using the maximum Q. The use of shared bandwidth reduction has a larger impact on performance, as a result of losing spatial information in part of the spectrum. This can be observed by comparing the curves with the same type of mark,

[2] http://home.tiscali.nl/ehabets/rir_generator.html

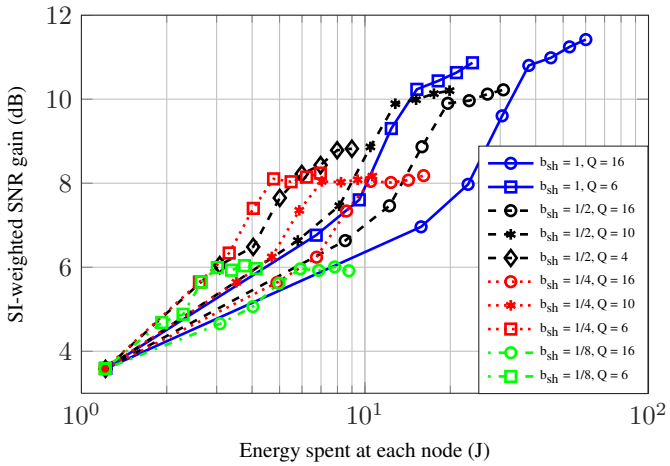

Fig. 2. Trade-off between energy and noise reduction performance in the simulated scenario.

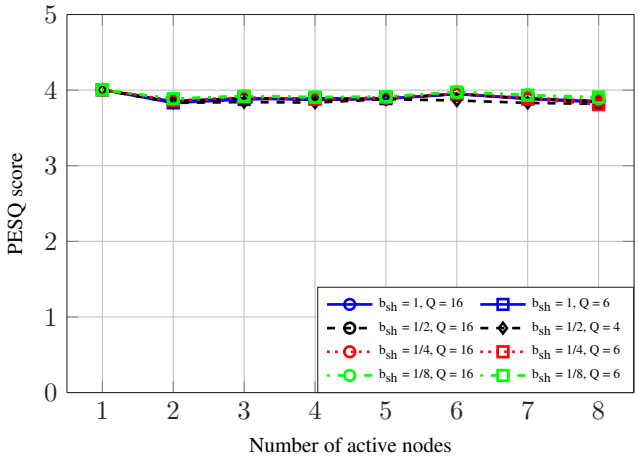

Fig. 3. PESQ scores of the output speech component for different operating parameters.

e.g. circle, where we observe that the energy savings are also larger, up to one eighth using shared bandwidth reduction with the maximum Q. The reason is that, although the communication cost is proportional to both L_{sh} and Q, L_{sh} can be reduced to a smaller fraction of its maximum value.

5.2. Speech distortion

To evaluate the speech distortion we chose the PESQ measure, an objective method which predicts the speech quality perceived by a human listener. Its goal is to compare the clean and degraded signals and give a score of the speech quality in a scale from 0 to 5 [10]. Since our interest is to analyze the distortions on the speech waveform, in our simulations we compare the input and output speech signals without noise. As shown in Fig. 3, the shared bandwidth reduction and the quantization do not significantly affect the speech distortion. The reason is that these modifications are only applied to the shared signals and not to the node's own signals. This is important because it shows that the energy consumption can be reduced at the expense of the noise reduction performance while having a small impact on the speech waveform.

6. CONCLUSIONS

We have studied energy-vs-performance trade-offs in the DANSE algorithm applied to speech enhancement for wireless acoustic sensor networks. We have proposed two algorithm modifications that allow nodes to spend less energy, at the cost of a reduction in the speech enhancement performance. Compared to the strategy of shutting down nodes, these modifications provide more flexibility to adjust the energy consumption and the desired performance. In order to analyze the energy spent by a node while executing the algorithm, we have provided an energy model that accounts for the energy consumed in computation and communication. Simulations have shown that our modifications allow nodes to significantly scale down their energy consumption depending on the tolerated reduction in performance. These results show significant potential for extending the network lifetime using dynamic system reconfiguration, which will be the topic of future work.

REFERENCES

[1] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537 – 568, 2009.

[2] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: Sequential node updating," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5277 –5291, oct. 2010.

[3] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part II: Simultaneous and asynchronous node updating," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5292 –5306, oct. 2010.

[4] A. Bertrand, J. Callebaut, and M. Moonen, "Adaptive distributed noise reduction for speech enhancement in wireless acoustic sensor networks," in *Proc. of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, August 2010.

[5] A. Bertrand and M. Moonen, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2196–2210, May 2011.

[6] P. Loizou, *Speech Enhancement: Theory and Practice*, CRC Press, 2007.

[7] F. Rosas and C. Oberli, "Modulation and SNR optimization for achieving energy-efficient communications over short-range fading channels," *IEEE Trans. on Wireless Communications*, vol. 11, no. 12, pp. 4286–4295, December 2012.

[8] Logos Electromechanical, "Zigduino homepage," 2015, http://www.logos-electro.com/store/zigduino-r2.

[9] ANSI S.3.5-1997, "American national standard methods for calculation of the speech intelligibility index," Tech. Rep., Acoust. Soc. America, June 1997.

[10] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Tech. Rep., ITU-T, February 2001.